

AD-A173 895

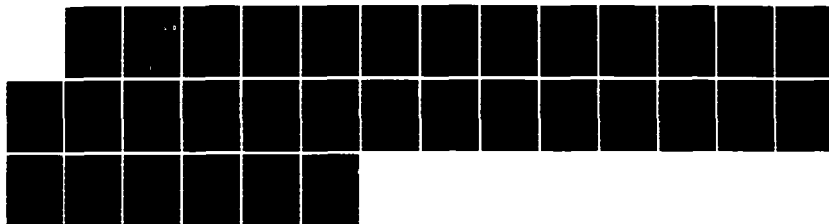
CONFIDENCE SETS FOR A CHANGE-POINT(U) STANFORD UNIV CA  
DEPT OF STATISTICS D SIEGMUND OCT 86 TR-39  
N00014-77-C-0306

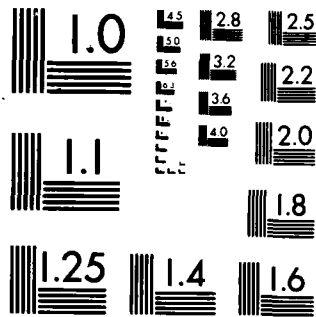
1/1

UNCLASSIFIED

F/G 12/1

NL





MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

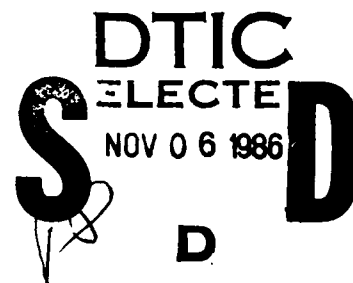
AD-A173 895

CONFIDENCE SETS FOR A CHANGE-POINT

by

David Siegmund  
Stanford University

TECHNICAL REPORT NO. 39  
OCTOBER 1986

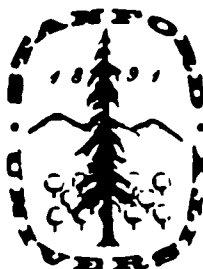


PREPARED UNDER CONTRACT  
N00014-77-C-0306 (NR-042-373)  
FOR THE OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted  
for any Purpose of the United States Government

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS  
STANFORD UNIVERSITY  
STANFORD, CALIFORNIA



DTIC FILE COPY

86 11 047

CONFIDENCE SETS FOR A CHANGE-POINT

by

David Siegmund  
Stanford University

TECHNICAL REPORT NO. 39

OCTOBER 1986

PREPARED UNDER CONTRACT

N00014-77-C-0306 (NR-042-373)

FOR THE OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted  
for any Purpose of the United States Government

Approved for public release; distribution unlimited

Also prepared under National Science Foundation Grant DMS86-00235 and issued as Technical  
Report No. 260, Department of Statistics, Stanford University.

DEPARTMENT OF STATISTICS  
STANFORD UNIVERSITY  
STANFORD, CALIFORNIA



Made for	
AMS	CRA&I
NO	TAB
Reviewed	
Distribution	
By	
Distribution /	
Availability Codes	
Not	Avail and/or Special
A-1	

# CONFIDENCE SETS FOR A CHANGE-POINT

David Siegmund

## Summary.

Several methods are discussed for confidence set estimation of a change-point in a sequence of independent observations from completely specified distributions. The method based on the likelihood ratio statistic is extended to the case of independent observations from a one parameter exponential family. Joint confidence sets for the change-point and the parameters of the exponential family are also considered.

## 1. Introduction.

Let  $x_1, x_2, \dots, x_m$  be independent random variables with  $x_1, \dots, x_j$  having distribution  $F$  and  $x_{j+1}, \dots, x_m$  having distribution  $G \neq F$ . The change-point  $j$ , where the distribution shifts from  $F$  to  $G$ , is an unknown parameter, to be estimated by a confidence set. In general, the distributions  $F$  and  $G$  may be known, completely unknown, or specified up to an unknown parameter. In this paper I discuss several procedures for the artificial but informative case of completely specified  $F$  and  $G$ , and then develop more completely a method based on the likelihood ratio statistic for the case where  $F$  and  $G$  come from a common one parameter exponential family of distributions. Precedent for the approach taken here is found in Worsley (1986) and Siegmund (1986).

Section 2 is concerned with known  $F$  and  $G$ . In addition it is assumed that the sequence of observations is actually doubly infinite,  $\dots, x_{-1}, x_0, x_1, \dots$ . This additional assumption has little effect if  $m$  is large and it is known that  $j$  is not close to 1 nor to  $m$ , because observations far from the change-point carry little information about the location of the change-point. The virtue of the assumption is that it makes  $j$  into a location parameter and provides an exact ancillary statistic: the class of shift invariant events. Five confidence set estimates are discussed. Three are studied by Siegmund (1986), in the context of estimating a change-point

in the drift of Brownian motion. The fourth is essentially the suggestion of Cobb (1978), and the fifth has smallest expected size among all shift invariant confidence sets. Section 3 compares the different confidence sets.

Sections 4 and 5 are concerned with the case that  $F$  and  $G$  are imbedded in a common one parameter exponential family, whose parameter  $\theta$  is unknown. Section 4 develops a method based on the likelihood ratio statistic for obtaining exact confidence sets for  $j$ . A new, fairly simple approximation is suggested for the required probability calculation. The approximation is illustrated on the coal mining accident data along the lines discussed by Worsley (1986). Section 5 involves the special case of normal distributions with  $j$  denoting a change in the mean. The likelihood ratio method is extended to give a joint confidence set for  $j$  and the difference between the two means.

## 2. The Cases of Known $F$ and $G$ .

Let  $\mathbb{Z}$  denote the integers and let  $j \in \mathbb{Z}$ . Let  $x_n, n \in \mathbb{Z}$  be a sequence of independent random variables with  $x_n$  having the distribution function  $F$  or  $G$  according as  $n \leq j$  or  $n > j$ . The distributions  $F$  and  $G$  are assumed known; the change-point  $j$  is unknown. Let  $P_j$  denote the probability measure induced by this model on the space of infinite sequences  $\omega = (x_n, n \in \mathbb{Z})$ . Let  $\sigma$  denote the shift operator, i.e., the mapping which takes  $\omega = (x_n, n \in \mathbb{Z})$  into  $\sigma\omega = (x_{n+1}, n \in \mathbb{Z})$ . Note that the family  $\{P_j, j \in \mathbb{Z}\}$  is a translation family in the sense that for any event  $B$  and  $j \in \mathbb{Z}$

$$P_j(B) = P_j(\omega \in B) = P_0(\sigma^{-j}\omega \in B) = P_0(\sigma^j B).$$

Let  $z_n = \log\{dG(x_n)/dF(x_n)\}$  denote the log likelihood ratio of  $x_n$ , and put

$$\begin{aligned}\tilde{S}_n &= z_1 + \dots + z_n & (n \geq 1) \\ &= -(z_{n+1} + \dots + z_0) & (n \leq -1) \\ &= 0 & (n = 0)\end{aligned}$$

Let  $\ell_i = dP_i/dP_0$  denote the likelihood function at  $i$ . By considering the finite sequence  $x_n, -N \leq n \leq N$ , and then letting  $N \rightarrow \infty$ , one can easily show that  $\ell_i = \exp(\tilde{S}_i)$ . Under  $P_0$

the log likelihood process  $(\tilde{S}_n, n \in \mathbb{Z})$  is a random walk satisfying  $\tilde{S}_0 = 0$  and having increments  $\tilde{S}_n - \tilde{S}_{n-1}$  with mean  $\int \log(dG/dF)dF < 0$  for  $n > 0$  and  $\int \log(dF/dG)dF > 0$  for  $n \leq 0$ .

The maximum likelihood estimator for  $j$  is the value  $\hat{j}$  where the process  $(\tilde{S}_n, n \in \mathbb{Z})$  assumes its maximum value. In general this value need not be unique, but to avoid technicalities it is assumed to be so in what follows. In the space of the sufficient statistic  $(\tilde{S}_n, n \in \mathbb{Z})$ , the sequence  $Y_i = \tilde{S}_{\hat{j}+i} - \tilde{S}_{\hat{j}}$ ,  $i \in \mathbb{Z}$ , is ancillary.

In the context of estimating a change-point in the drift of a Brownian motion process, Siegmund (1986) compares the following three confidence sets for the change-point  $j$ . The first two were discussed earlier by Hinkley (1970, 1972), who, however, made no attempt to establish their relative efficiency.

(i) Since  $\hat{j} - j$  is pivotal, if  $r = r_\alpha$  is defined by  $P_0(|\hat{j}| > r) = \alpha$ , then  $C_1 = [\hat{j} - r, \hat{j} + r]$  is a  $(1 - \alpha)$  100% confidence interval.

(ii) Let  $A_j$  denote the acceptance region of a size  $\alpha$  likelihood ratio test of the hypothesis that the change-point is  $j$ , i.e.,  $A_j = \{\max_n \tilde{S}_n - \tilde{S}_j < \eta\}$ , where  $\eta = \eta_\alpha$  satisfies  $P_j(A_j) = \{P_0(\max_{n \geq 0} \tilde{S}_n < \eta)\}^2 = 1 - \alpha$ . Then the set  $C_2$  of  $n \in \mathbb{Z}$  such that the observed sample point  $\omega \in A_n$  is a  $(1 - \alpha)$  100% confidence set. Since the log likelihood process  $(\tilde{S}_n, n \in \mathbb{Z})$  is in general multimodal, this confidence set is not in general an interval.

(iii) A modification of the preceding method which always yields an interval is to define

$$L(R) = \min(\max) \left\{ n : \tilde{S}_n \geq \max_i \tilde{S}_i - \eta' \right\},$$

which for suitable  $\eta' < \eta$  satisfies

$$P_j(L \leq j \leq R) = P_0(L \leq 0 \leq R) = 1 - 2P_0(R < 0) = 1 - \alpha.$$

The next possibility is essentially the suggestion of Cobb (1978). In analogy with Fisher's (1934) observation that the conditional probability density of the maximum likelihood estimator of a location parameter given the sample spacings, which are ancillary in that case, is the normalized likelihood function, one may show by a direct calculation that

$$P_j(\hat{j} - j = n | Y_i, i \in \mathbb{Z}) = P_0(\hat{j} = n | Y_i, i \in \mathbb{Z}) = \exp(\tilde{S}_{\hat{j}_{obs} - n}) / \sum_i \exp(\tilde{S}_i), \quad (1)$$

where  $\hat{j}_{obs}$  denotes the observed value of  $\hat{j}$ . Let

$$p_n = \exp(\tilde{S}_n) / \sum \exp(\tilde{S}_i), \quad n \in \mathbb{Z}. \quad (2)$$

(iv) It follows from (1) that a confidence set of conditional coverage probability  $1 - \alpha$  can be formed as follows. Order the  $p_n$  in (2) as  $p_{(1)} \geq p_{(2)} \geq \dots$ . Construct the set  $C_4$  by putting the index  $n_1$  corresponding to  $p_{(1)}$  in  $C_4$  and continuing to add points  $n_2, \dots, n_k$  corresponding to  $p_{(2)}, \dots, p_{(k)}$  as long as  $\sum_{i \leq k} p_{(i)} < 1 - \alpha$ . Note that for a Bayesian with a uniform prior on  $\mathbb{Z}$ ,

$$p_n = \text{pr}(j = n | x_i, i \in \mathbb{Z})$$

and hence the set  $C_4$  is a highest posterior probability credible set for  $j$ . In fact, even without the explicit evaluation in (1), one knows from a general theorem of Stein (1965) and Hora and Buehler (1966) that the highest posterior credible set for  $j$  is also a confidence set.

(v) One can also obtain an unconditional confidence set from the formal posterior probabilities  $(p_n, n \in \mathbb{Z})$  in (2) as follows: let  $c$  be such that

$$P_j\{p_j \geq c\} = P_0\left\{\sum \exp(\tilde{S}_n) \leq c^{-1}\right\} = 1 - \alpha, \quad (3)$$

and  $C_5 = \{n : p_n \geq c\}$ . Then  $C_5$  is a  $(1 - \alpha)$  100% confidence set, which according to a general theorem of Hooper (1982) or alternatively by a simple Neyman-Pearson argument has smallest expected size among all shift equivariant confidence sets.

**Remarks.** The confidence sets (ii), (iv), and (v) all order the parameter values for inclusion according to the value of the likelihood function. Where they disagree is where to draw the line between inclusion and exclusion. For those who strongly prefer a confidence interval to a



possibly disconnected confidence set, (iii) appears to be a reasonable modification of (ii). It is possible to give analogous modifications of (iv) and (v).

Of these five confidence sets, all except for (iv) require computation of a sampling distribution. Approximations are suggested in the following section.

### 3. Comparisons.

The purpose of this section is to compare the expected size of the various confidence sets proposed in Section 2. Since the case of known  $G$  and  $F$  is artificially simple and our main goal is insight into the case where  $G$  and  $F$  contain unknown nuisance parameters, there seems to be little harm in simplifying the technical problems somewhat by assuming that  $F$  is  $N(0, 1)$  and  $G$  is  $N(\delta, 1)$  for a known  $\delta > 0$ .

Siegmund (1986) considers the computationally simpler case of a Brownian motion process and shows that the length of the confidence interval defined in (i) is substantially longer than the expected size of the confidence sets in (ii) and (iii).

In the present context it can be shown as  $\alpha \rightarrow 0$  that the expected sizes of the confidence sets in (ii) - (v) are all  $\sim 4\delta^{-2} \log \alpha^{-1}$ , whereas the length of the interval in (i) is  $\sim 8\delta^{-2} \log \alpha^{-1}$ . Hence the confidence interval  $C_1$  defined in (i) appears not to be competitive with the others and will not be considered further.

Although Siegmund's (1986) comparison of (ii) and (iii) favors (ii), the difference is not large. In fact there is a transcription error in passing from the first to the second line of the display following (3.15) of Siegmund (1986), and consequently the difference in the numerical example between methods (ii) and (iii) is smaller than stated there. Since one suspects that the rapid fluctuations of Brownian motion may account for some of that difference, and since (iii) is the only remaining interval estimate and is a surrogate for interval modifications of (iv) and (v), it seems reasonable to make a comparison of (ii) and (iii) in the present discrete time setting. Theorem 1 below gives asymptotic expansions as  $\alpha \rightarrow 0$  of the expected size of the confidence sets (ii) and (iii).

It seems difficult to give comparably precise expansions for (iv) and (v). Hence (ii),

(iv), and (v) are compared below in a Monte Carlo experiment, which also shows that the approximations given in Theorem 1 are reasonably accurate.

We begin with approximations for the coverage probability of (ii) and (iii). Let  $\Phi$  be the standard normal distribution function and

$$\nu(x) = 2x^{-2} \exp \left\{ -2 \sum_{n=1}^{\infty} n^{-1} \Phi \left( -x\sqrt{n}/2 \right) \right\} \quad (x > 0). \quad (4)$$

For computational purposes it usually suffices to use the small  $x$  approximation (Siegmund, 1985, p. 219)

$$\nu(x) = \exp(-\rho x) + o(x^2) \quad (x \rightarrow 0), \quad (5)$$

where  $\rho \cong .583$ . For the normally distributed  $x_n, n \in \mathbb{Z}$ , under consideration here  $\tilde{S}_n = \delta(n\delta/2 - S_n), n = 0, 1, \dots$ , where  $S_n = x_1 + \dots + x_n$ . It follows from a classical result of Cramér (cf. Siegmund, 1985, (8.49)) that

$$P_0 \left( \max_{n \geq 0} \tilde{S}_n \geq \eta \right) \sim \nu(\delta) \exp(-\eta) \quad (\eta \rightarrow \infty) \quad (6)$$

and hence by (5) for  $A_j$  defined in (ii) above

$$P_j(A_j) \cong \{1 - \exp(-\eta - \rho\delta)\}^2. \quad (7)$$

By conditioning on  $\max_{n \geq 0} \tilde{S}_n$ , one may show for  $R$  defined in (iii),

$$\begin{aligned} P_0(R < 0) &= P_0 \left( \max_{n \leq 0} \tilde{S}_n > \max_{n \geq 0} \tilde{S}_n + \eta' \right) \\ &\sim \nu(\delta) \exp(-\eta') E_0 \left\{ \exp \left( - \max_{n \geq 0} \tilde{S}_n \right) \right\} \end{aligned} \quad (8)$$

$\eta' \rightarrow \infty$ . It is possible to compute the expectation on the right hand side of (8) numerically or give a small  $\delta$  expansion analogous to (5), but for our purposes it seems adequate to pretend that (6) is an equality, which leads to the approximation

$$P_j(0 \notin [L, R]) \cong 2 \exp(-\eta' - \rho\delta) \{1 - \exp(-\rho\delta)/2\}. \quad (9)$$

The following theorem gives an asymptotic expansion as  $\alpha \rightarrow 0$  of the expected size of  $C_2$  defined in (ii) and  $[L, R]$  defined in (iii). It will be convenient to use the notation  $[y] =$  integer part of  $y$ ,  $|C| =$  number of elements in the set  $C$ , and  $M = \sup_{n \geq 0} \tilde{S}_n$ .

**Theorem 1.** Let  $C_2$  be the confidence set defined in (ii) and  $[L, R]$  the confidence interval defined in (iii). As  $\eta \rightarrow \infty$

$$E_j|C_2| = 2[2\eta/\delta^2] + 4/\delta^2 - 4\delta^{-1} \int_0^\infty \{2P_0(M > x) - P_0^2(M > x)\} dx + o(1),$$

and as  $\eta' \rightarrow \infty$

$$E_j(R - L) = 2[2\eta'/\delta^2] + 4\delta^2 - 4\delta^{-1} \int_0^\infty \int_0^\infty P_0(M \in dy) \{2P_0(M > x + y) - P_0^2(M > x + y)\} dx + o(1).$$

A proof is sketched in an appendix.

To obtain easily evaluated approximations to the integrals appearing in these expressions, one may again pretend that (6) is an equality and use (5). This leads to

$$E_j|C_2| \cong 2[2\eta/\delta^2] + 2\delta^{-2} (2 - 4e^{-\rho\delta} + e^{-2\rho\delta}) \quad (10)$$

and

$$E_j(R - L) \cong 2[2\eta'/\delta^2] + 2\delta^{-2} (2 - 4e^{-\rho\delta} + 3e^{-2\rho\delta} - 2e^{-3\rho\delta}/3). \quad (11)$$

Table 1 contains some numerical examples. It indicates that there is essentially no difference between the expected size of the confidence sets (ii) and (iii). On the basis of these

results a statistician who strongly prefers a confidence interval to the generally disconnected likelihood ratio confidence set should feel comfortable in imposing that constraint.

Table 1.  
Expected Size of Confidence Sets (ii) and (iii)

$\alpha$	$\delta$	$\eta$ (7)	$E_0 C_2 $ (10)	$\eta'$ (9)	$E_0(R - L)$ (11)
.1	0.7	2.56	19.1	2.18	17.9
.1	1.0	2.39	8.2	2.08	9.2
.05	0.7	3.27	25.1	2.88	23.9
.05	1.0	3.09	12.2	2.78	11.2
.01	0.7	4.89	37.1	4.49	37.9
.01	1.0	4.71	18.2	4.39	17.2

In the present context of completely specified distributions there is no sampling theory to develop in order to use the confidence set (iv). However, it seems a difficult problem to give a reasonable approximation for the related set defined in (v). A crude approximation to (3) which might be used as the first step in an iterative numerical or Monte Carlo scheme is to replace  $\tilde{S}_n$  by a Brownian motion process  $\tilde{W}(t)$  with drift  $-(\delta^2/2)\text{sgn}(t)$  and variance  $\delta^2$  and replace the sum in (3) by an integral. One easily sees that the integral over  $[0, \infty)$  has the distribution given by Pollak and Siegmund (1985, Proposition 3). This can be convolved with itself to obtain  $\text{pr}[\int_{-\infty}^{\infty} \exp\{\tilde{W}(t)\}dt < c^{-1}] = 2\delta^{-1}\sqrt{c} \exp(-4c/\delta^2)K_1(2\delta^{-1}\sqrt{c})$ , where  $K_1$  is the modified Bessel function of the second kind.

Table 2 reports the results of 1000 repetition Monte Carlo experiment with  $m = 100$  and  $j = 50$  to compare the confidence sets  $C_2, C_4$ , and  $C_5$ . It confirms that the analytic approximation for the expected size of  $C_2$  given in Theorem 1 is reasonably accurate and shows that all three confidence sets have about the same expected size.

**Table 2.**  
**Monte Carlo Comparison of  $C_2$ ,  $C_4$ , and  $C_5$**

$\alpha$ (nominal)	$C_2$			$C_4$		$C_5$		
	$\delta$	$\hat{\alpha}$	$E_0 C_2 $	$\hat{\alpha}$	$E_0 C_4 $	$c$	$\hat{\alpha}$	$E_0 C_5 $
.10	.07	.090	18.8	.084	19.5	.010	.092	19.3
.10	1.0	.098	9.6	.085	10.3	.022	.113	9.4
.05	0.7	.041	24.6	.040	25.2	.005	.047	26.0
.05	1.0	.048	12.6	.037	13.2	.011	.052	12.6

Although the confidence sets defined in (ii)-(iv) perform similarly on the average, they can treat individual sets of data differently. Figure 1 displays two simulated log likelihoods with  $m = 101$ ,  $j = 50$ , and  $\delta = 0.7$ . The horizontal line defines the 95% likelihood ratio confidence set (ii). In accordance with the approximation (7) it is drawn 3.27 units below the maximum of the log likelihood function.

In the upper part of Figure 1 the one major peak of the log likelihood is fairly sharp with the consequence that all the confidence sets are about one half their expected size of 25. The confidence interval defined in (iii) has one point less on each end than the likelihood ratio confidence set. The formal Bayes posterior set,  $C_4$ , makes a smaller adaptation to the peaked log likelihood; it contains four more points, including the local maximum at 63. The confidence set  $C_5$  is the same as the likelihood ratio confidence set.

The lower part of Figure 1 contains a comparatively flat log likelihood with two distinct peaks. The likelihood ratio confidence set contains 33 points. The interval modification is now slightly larger because it contains points of relatively low likelihood: 44, 45, 56-58. Again the formal Bayes posterior set adapts less to the departure of the log likelihood from its expected shape and this time contains four fewer points than the likelihood ratio confidence set.

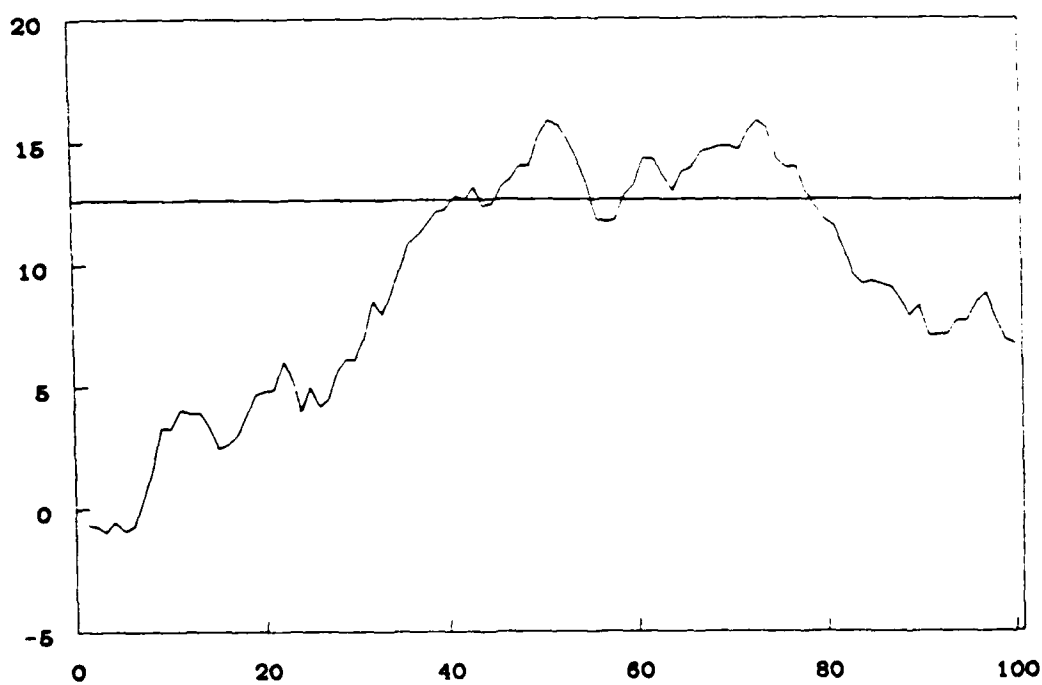
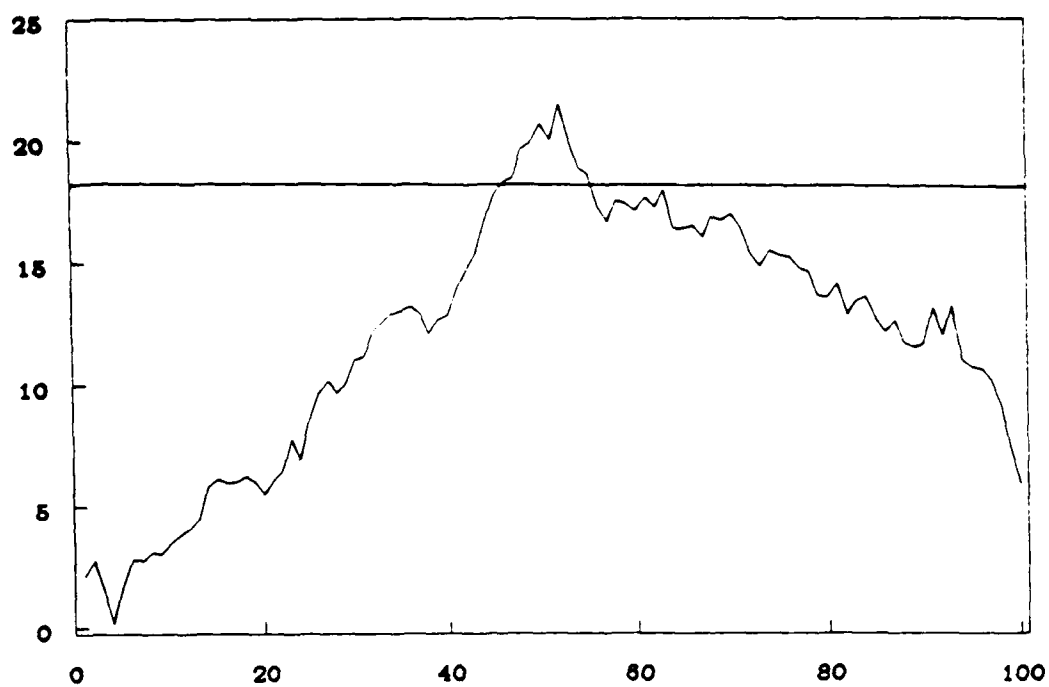


Figure 1. Two Simulated Log Likelihoods.

In general, the interval modification (iii) is usually slightly shorter than the likelihood ratio confidence set but can be considerably larger. The formal Bayes posterior set is usually larger than the likelihood ratio when both sets are small and smaller when both sets are large. This suggests that there may be recognizable subsets making the conditional coverage probability of the likelihood ratio set differ from its nominal value. The confidence set  $C_5$  can look rather foolish conditionally. If all the  $p_i$  are very small and about equal, it can deliver a small, or perhaps empty confidence set while the other methods recognize the data as uninformative and yield large confidence sets. Presumably this occurs with small probability.

Overall the evidence given here does not seem persuasive for choosing among the confidence sets (ii) - (v). A possible conclusion is that in more complex problems one may reasonably use whichever method seems most easily adapted to the problem at hand. When the distributions  $F$  and  $G$  are unknown, but can be imbedded in a common exponential family, one can use a conditioning argument to obtain exact likelihood ratio confidence sets. This is the subject of the next section.

#### 4. The Likelihood Ratio Method for an Exponential Family.

Now suppose that  $F$  and  $G$  can be imbedded in an exponential family of the form

$$dF_\theta(x) = \exp\{\theta x - \psi(\theta)\} dF_0(x)$$

relative to some fixed distribution  $F_0$ , which without loss of generality can be standardized to have mean 0 and variance 1. Thus for some unknown  $\theta_0 \neq \theta_1$  and  $j \in \{1, \dots, m\}$ ,  $x_1, \dots, x_j$  have distribution  $F_{\theta_0}$  and  $x_{j+1}, \dots, x_m$  have distribution  $F_{\theta_1}$ . The probability on the space of  $x_1, \dots, x_m$  will be denoted by  $\text{pr}$ , with the dependence on  $j, \theta_0$ , and  $\theta_1$  suppressed.

Several writers, e.g., Davies (1977), Siegmund (1986), and Worsley (1986), have observed that one can extend the likelihood ratio method (ii) of Section 2 to obtain a confidence set for  $j$  in the presence of the unknown nuisance parameters  $\theta_0, \theta_1$  as follows. Let  $H(x) = \sup_\theta \{\theta x - \psi(\theta)\}$ ,  $S_n = x_1 + \dots + x_n$ , and

$$\Lambda_{n,m} = nH(n^{-1}S_n) + (m-n)H\{(m-n)^{-1}(S_m - S_n)\}. \quad (12)$$

The likelihood ratio test of the hypothesis that the change-point is  $j$  has acceptance region of the form

$$A_j = \left( \max_n \Lambda_{n,m} - \Lambda_{j,m} \leq k \right).$$

By sufficiency the conditional probability of  $A_j$  given  $(S_j, S_m)$  does not depend on  $\theta_0, \theta_1$ . Hence if one chooses  $k = k(\xi_1, \xi_2)$  so that

$$\text{pr}(A_j | S_j = \xi_1, S_m = \xi_2) = 1 - \alpha$$

for all  $\xi_1, \xi_2$ , then the set of values  $j$  which are accepted by the test is a  $(1 - \alpha)100\%$  confidence set.

It is not actually necessary to solve for  $k(\xi_1, \xi_2)$  in order to determine the confidence set. Given  $S_j$  and  $S_m$ ,  $\Lambda_{j,m}$  is constant, and hence the confidence set is most easily determined as the set of  $j$  for which

$$\text{pr} \left\{ \max_n \Lambda_{n,m} \leq (\max_n \Lambda_{n,m})_{obs} | S_j, S_m \right\} \leq 1 - \alpha. \quad (13)$$

An approximation for this conditional probability which seems adequate for many cases is given below.

Note that one might also define  $A_j$  as the acceptance region of the likelihood ratio test in the conditional model given  $S_m$ . The unconditional test is often simpler analytically, but the conditional one may turn out to be preferable. In the simplest case of a normal distribution with mean  $\theta$  and variance 1 there is no difference between the tests. Similarly one could substitute Pettitt's (1980) test with acceptance region of the form

$$A_j = \left\{ \max_n |(nS_m/m - S_n)| - |(jS_m/m - S_j)| \leq a \right\}. \quad (14)$$

Given  $(S_j, S_m)$  the random variables  $\max_{n \leq j} \Lambda_{n,m}$  and  $\max_{j \leq n < m} \Lambda_{n,m}$  are conditionally independent, and hence the left hand side of (13) is of the form



$$\text{pr} \left( \max_{n \leq j} \Lambda_{n,m} \leq a | S_j, S_m \right) \text{pr} \left( \max_{j \leq n < m} \Lambda_{n,m} \leq a | S_j, S_m \right). \quad (15)$$

These two probabilities present similar computational problems, so it suffices to consider the second one, or equivalently

$$\text{pr} \left( \max_{j \leq n < m} \Lambda_{n,m} > a | S_j, S_m \right). \quad (16)$$

In the special case that  $F_\theta$  is the normal distribution with mean  $\theta$  and variance 1, the probability (16) equals

$$\text{pr} \left\{ \max_{j \leq n < m} \frac{(nS_m/m - S_n)^2}{2n(1 - n/m)} > a | jS_m/m - S_j = \xi \right\}, \quad (17)$$

for which Siegmund (1986) gives an approximation under the assumptions that  $j, a$ , and  $\xi$  are all proportional to  $m$ , and

$$c^2 = 2a - \xi^2 / \{j(1 - j/m)\} \quad (18)$$

is asymptotically a positive multiple of  $m$  as  $m \rightarrow \infty$ . A somewhat simpler approximation is obtained by assuming that  $c^2 = o(m)$ . For ease of reference we record the result as Theorem 2.

**Theorem 2.** Let  $x_1, \dots, x_m$  be independent standard normal random variables and  $S_n = x_1 + \dots + x_n$ . Let  $\min(j, m - j), a$ , and  $\xi$  be proportional to  $m$  as  $m \rightarrow \infty$ . Suppose  $c^2$  defined by (18) diverges to  $+\infty$  but  $c^2 = o(m)$ . Then the probability (17) is

$$\sim \nu[\xi / \{j(1 - j/m)\}] \exp(-c^2/2) \quad (19)$$

as  $m \rightarrow \infty$ , where  $\nu$  is defined in (4) and given approximately by (5).

From the simulations reported in Table 6 of Siegmund (1986) one can see that (19) is reasonably accurate for the range of  $j, m$ , and  $\xi$  considered there. Presumably it is less accurate for larger  $c$  and/or smaller  $m$ , but it seems adequate for many cases of interest.

According to the approximations (19) and (5) the confidence set defined by (13) is the set of all  $i$  such that

$$\left\{ 1 - \exp\left(-.583[2\Lambda_{i,m}/\{i(1-i/m)\}]^{1/2} - (\max_n \Lambda_{n,m} - \Lambda_{i,m})\right) \right\}^2 \leq 1 - \alpha. \quad (20)$$

Even when one questions the accuracy of (19) or when the data are not normal, the central limit theorem suggests the use of (20) as a first approximation. A better approximation, simulation, or numerical methods can be used to decide whether values of  $i$  on the borderline according to (20) should be included in or excluded from the confidence set.

Note also the formal similarity between (19) and (6). To the extent that  $\{i(1-i/m)\}^{1/2}$  is nearly constant over the values  $i$  of interest, e.g., when the likelihood ratio statistic is sharply peaked and hence the confidence set is small, (20) shows that the confidence set consists of those  $i$  for which  $\Lambda_{i,m}$  is within some distance of  $\max_n \Lambda_{n,m}$ , which can be displayed graphically as in Section 2.

Figure 2 shows the log likelihood ratio statistic and the approximate cutoff for a 95% confidence set for the same simulated data as in Figure 1. Qualitatively the cases of known and unknown  $\delta$  look quite similar. Usually the confidence set is larger in the case of unknown  $\delta$ , and this is indeed so in the lower plot. However, the reverse is true in the upper plot, presumably because the procedure in effect estimates  $\delta$  and then acts as if the, in this case large, estimated value is the true one.

Returning to the general exponential family, if we let  $a = ma_0$  and condition on  $S_m = \xi_2 = m\xi_{20}$ , we see from (12) that  $(\max_{j \leq n < m} \Lambda_{n,m} > a) = \bigcup_{n=j}^{m-1} [\{S_n > mb_2(n/m)\} \cup \{S_n < mb_1(n/m)\}]$ ,

where  $b_1(t) < b_2(t)$ ,  $0 < t < 1$ , are the solutions of

$$tH\{t^{-1}b_i(t)\} + (1-t)H[(1-t)^{-1}\{\xi_{20} - b_i(t)\}] = a_0. \quad (21)$$

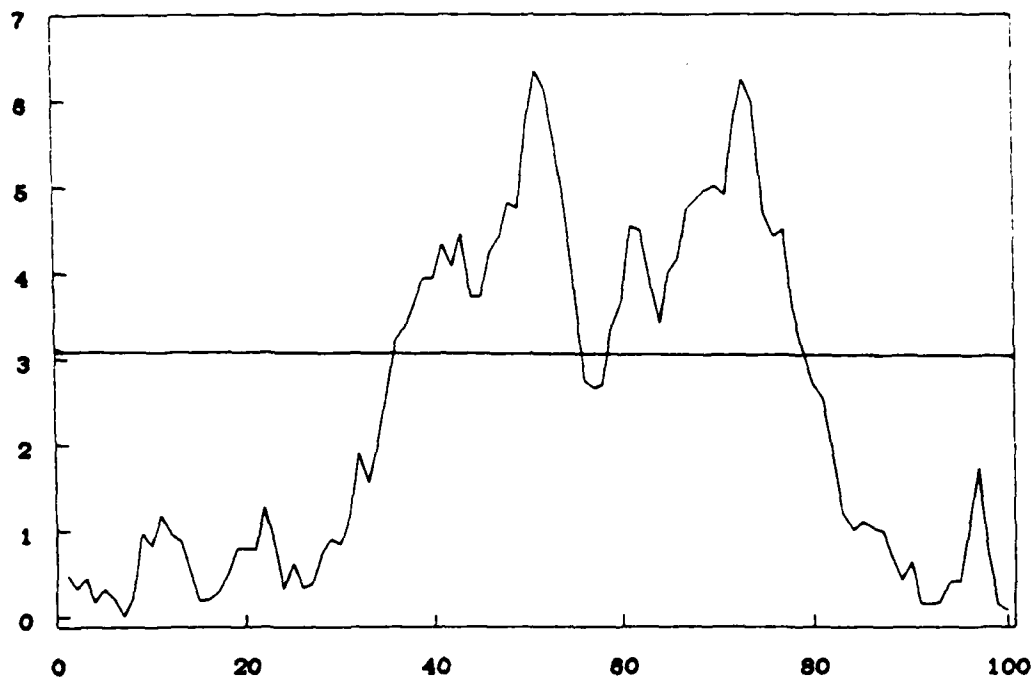
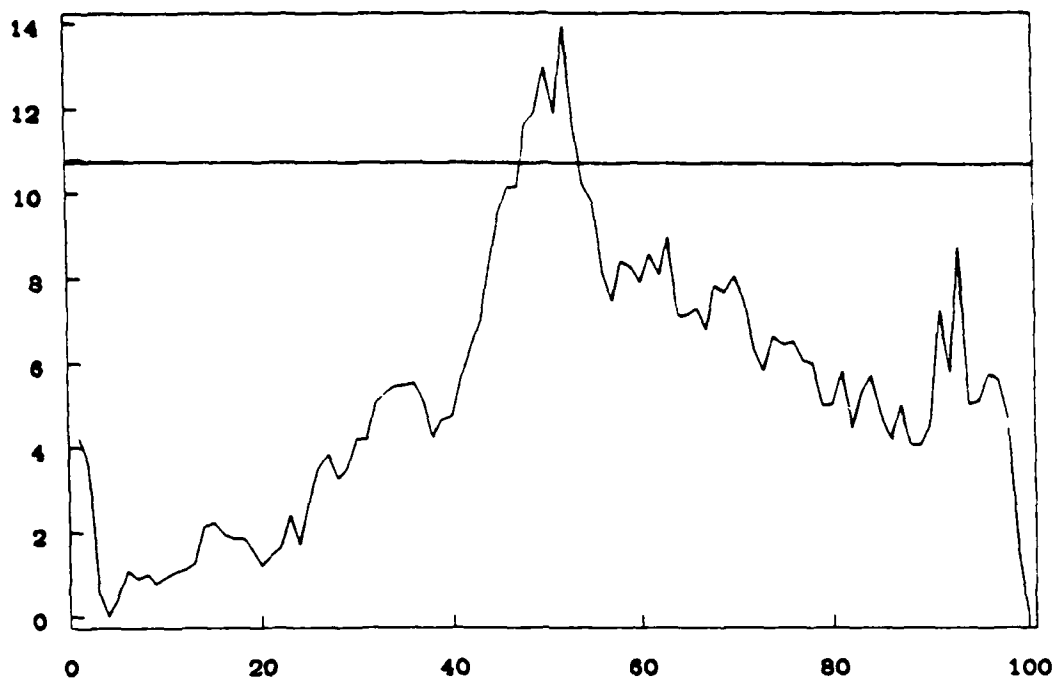


Figure 2. Simulated Log Likelihood Ratio Processes,  $\delta$  Unknown.

Usually one is interested in evaluating (16) in cases where  $S_j = \xi_1$  is fairly close to one of the boundary curves  $mb_1(j/m)$  or  $mb_2(j/m)$ . Thus the probability of crossing the other can be neglected, and it seems reasonable to develop an approximation in which the distance from  $\xi_1$  to the relevant curve is small in some sense. See Figure 3. Our problem reduces to approximate evaluation of probabilities like

$$\text{pr}[S_{j+i} > mb_2\{(j+i)/m\} \text{ for some } i < m-j | S_j = \xi_1, S_m = \xi_2]. \quad (22)$$

The mathematically convenient interpretation of the condition that  $mb_2(j/m) - \xi_1$  be small is that it be  $O(\sqrt{m})$ .

Siegmund (1985,1986) develops a method for approximating boundary crossing probabilities which can be adapted to the present context. A suitable result is given in Appendix B.

*As an illustration we consider the British coal mining accident data of Maguire, Pearson, and Wynn (1952), as extended and corrected by Jarrett (1979). Worsley (1986) has analysed the original data and determined the likelihood ratio confidence set by numerical computation of (15).*

The data are intervals in days between accidents in British coal mines in which at least ten deaths occurred. Jarrett's (1979) data involve  $m = 190$  intervals from 15 March, 1851 to 22 March, 1962, a period of 40,549 days. Under the assumption that the intervals  $y_1, \dots, y_m$  are independent and exponentially distributed with a change after the  $j$ -th observation in the mean time between accidents, we shall determine a likelihood ratio confidence set for  $j$ .

The likelihood ratio statistic is  $\max_n \Lambda_{n,m} = \max_n [m \log(W_m/m) - n \log(W_n/n) - (m-n) \log\{(W_m - W_n)/(m-n)\}]$ , where  $W_n = y_1 + \dots + y_n$ . For Jarrett's data the maximum value equals 35.6 and is assumed at  $n = 124$  in the year 1890. the approximation (20) gives the set  $\{116, 117, \dots, 128, 133\}$  as a 95% confidence set for the change-point. This corresponds to the interval from 1887 to 1893 together with an isolated point in 1897. One may want to use the presumably more accurate probability approximation given in Theorem 3 in Appendix B to check some of the borderline cases.

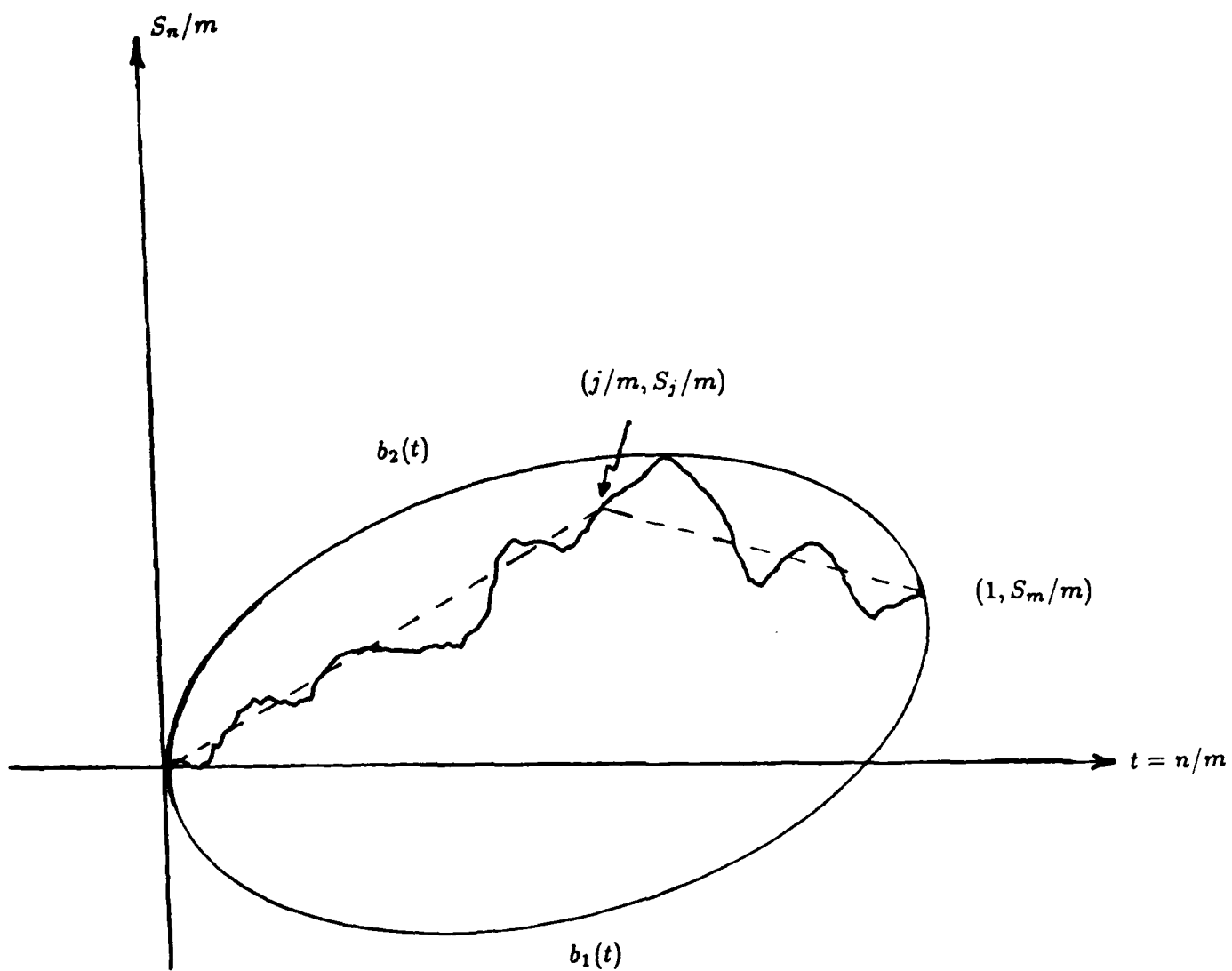


Figure 3. Conditional Boundary Crossing Problem.

For example, for  $j = 129$ ,  $\Lambda_{m,j} = 32.4$ , so (20) yields .961, and 129 is not included in the confidence set. From (21) with  $a = ma_0 = 35.6$  one easily calculates the ingredients to apply Theorem 3 in Appendix B and obtains as an approximation to (15)  $(1 - .011)(1 - .024) = .965$ , which confirms that 129 should be excluded from the confidence set. Note that this approximation and the normal approximation, (20), are reasonably consistent, although the normal approximation involves two equal factors while this one contains two unequal factors, one smaller and one larger than in the normal approximation. For  $j = 128$  the approximation by means of Theorem 3 for the second factor in (15) is  $1 - .057 = .943$ , and hence 128 is included in the confidence set regardless of the value of the first factor. After examining two or three values of  $i$ , one quickly concludes that the approximation of Theorem 3 yields the same confidence set as the crude normal approximation.

Application of (20) to the original Maguire, Pearson, and Wynn (1952) data gives precisely the same confidence set which Worsley computed numerically. However, because of discrepancies between the two data sets, the years covered by the two confidence sets are slightly different.

Raferty and Akman (1986) give a flat prior Bayesian analysis of these data. It appears from their calculations and Figure that a highest posterior set estimate for the change-point is the same as the confidence set computed here. Presumably such a postestrior set is, under some general conditions, approximately a confidence set for large  $m$ , but the elegant exact relation of (1) and (2) is no longer valid. It would be interesting to give some precise asymptotic results, which would serve to extend the method (iv) in Section 2 to the case of unknown nuisance parameters.

Cobb (1978) has suggested an alternative extension of method (iv) to deal with nuisance parameters, but it contains some arbitrary features which may make it difficult to implement with small or moderate sample sizes.

## 5. Joint Confidence Sets.

The likelihood ratio method can also be adapted to give joint confidence sets for the change-point  $j$  and some function  $\delta$  of the parameters  $\theta_0$  and  $\theta_1$ . In this section we consider

the simple case of normally distributed  $x_i$  having mean  $\theta_0$  or  $\theta_1$  according as  $1 \leq i \leq j$  or  $j < i \leq m$  and variance one, and take  $\delta = \theta_1 - \theta_0$ .

The acceptance region of the likelihood ratio test of the hypothesis that the parameters are  $j$  and  $\delta$  is

$$A_{j,\delta} = [\sup_i \Lambda_{i,m} - \delta \{jS_m/m - S_j - j(1 - j/m)/2\} \leq c^2/2],$$

where  $\Lambda_{i,m} = (iS_m/m - S_i)^2 / \{2i(1 - i/m)\}$  and  $c = c(j, \delta)$  is chosen to satisfy

$$\text{pr}(A_{j,\delta}) = 1 - \alpha$$

for all  $j$  and  $\delta$ . Note that

$$\begin{aligned} & \sup_i \Lambda_{i,m} - \delta \{jS_m/m - S_j - 2^{-1}\delta j(1 - j/m)\} \\ &= \sup_i \Lambda_{i,m} - \Lambda_{j,m} + \frac{\{jS_m/m - S_j - \delta j(1 - j/m)\}^2}{2j(1 - j/m)}; \end{aligned} \quad (23)$$

and since the first difference on the right hand side is necessarily non-negative, one obtains

$$\begin{aligned} \text{pr}(A_{j,\delta}^c) &= \text{pr}[|jS_m/m - S_j - \delta j(1 - j/m)| > c\{j(1 - j/m)\}^{1/2}] \\ &+ E[\text{pr}(A_{j,\delta}^c | jS_m/m - S_j); |jS_m/m - S_j - \delta j(1 - j/m)| \leq c\{j(1 - j/m)\}^{1/2}]. \end{aligned} \quad (24)$$

The first term on the right hand side of (24) is exactly  $2\{1 - \Phi(c)\}$ . According to Theorem 2

$$\begin{aligned} & \text{pr}(A_{j,\delta}^c | jS_m/m - S_j = \xi) \\ & \sim 2\nu[\xi/\{j(1 - j/m)\}] \exp \left[ -c^2/2 + \frac{\{\xi - \delta j(1 - j/m)\}^2}{2j(1 - j/m)} \right] \end{aligned}$$

provided the exponent diverges more slowly than  $m$  as  $m \rightarrow \infty$ . Substitution of this approximation into (24) yields

$$\begin{aligned} \text{pr}(A_{j,\delta}^c) &\cong 2\{1 - \Phi(c)\} + 2\varphi(c) \int_{-c}^c \nu[\delta + z/\{j(1 - j/m)\}^{1/2}] dz \\ &\cong 2\{1 - \Phi(c)\} + 4\nu(\delta)c\varphi(c), \end{aligned} \quad (25)$$

if we assume that  $c/\{j(1 - j/m)\}^{1/2}$  is small. Here  $\varphi$  and  $\Phi$  are the standard normal density and distribution function respectively.

Using (25) one can easily find an approximate confidence set by trial and error. Given  $i$ , one sets  $\delta$  equal to the estimator  $\hat{\delta}_i = (iS_m/m - S_i)/\{i(1 - i/m)\}$  and finds the value of  $c$  for which (25) equals  $\alpha$ . Thus by (23) one finds whether that  $i$  and some  $\delta$  are in the confidence set. Then one iteratively finds upper and lower bounds on  $\delta$  for that particular value of  $i$ . In principle this must be repeated for each  $i$ .

An extension of this method to non-normal exponential families requires consideration of special cases. The generalization of (23), in almost obvious notation, is

$$\begin{aligned} & \sup_i \Lambda_{i,m} - \Lambda_{j,m}(\delta) \\ &= \left( \sup_i \Lambda_{i,m} - \Lambda_{j,m} \right) + \{ \Lambda_{j,m} - \Lambda_{j,m}(\delta) \}. \end{aligned}$$

If  $\delta$  is a function of the difference in the natural parameters of the exponential family, e.g. if the parent populations are Poisson and  $\delta$  is the ratio of their means, by computing probabilities conditionally given  $S_m$ , one obtains a statistic whose distribution is parameterized by  $\delta$ . On the other hand, if the parent distributions are exponential and  $\delta$  is again the ratio of their means, because of invariance of the two sample problem under common changes of scale, the unconditional model and unconditional distribution of  $\Lambda_{j,m} - \Lambda_{j,m}(\delta)$  are appropriate.

The sampling theory seems rather complicated. Presumably a normal approximation using (25) suffices when  $m$  is large enough, but this needs to be investigated.

### Acknowledgement

I would like to thank M. Pollak and T. Sellke for very helpful discussions, I. Einott for some programming assistance, and Qiwei Yao for spotting the transcription error in Siegmund (1986), which is mentioned in Section 2.



## APPENDIX A

### Informal Proof of Theorem 1.

We consider only the confidence interval  $[L, R]$ . The proof for the likelihood ratio confidence set is similar and somewhat simpler. Since the confidence set is equivariant, it suffices to consider the case  $j = 0$ . To simplify the notation we shall write  $\text{pr}$  and  $E$  instead of  $P_0$  and  $E_0$ ,  $\eta$  instead of  $\eta'$ ,  $S_n$  instead of  $\tilde{S}_n$ , and take  $\delta = 1$ . Recall that  $M = \sup_{n \geq 0} S_n$ .

For arbitrary  $n_0 = 1, 2, \dots$

$$\begin{aligned}
 E(R - L) &= \sum_{-\infty}^{\infty} \text{pr}(L \leq n \leq R) = \text{pr}(L \leq 0 \leq R) + 2 \sum_1^{\infty} \text{pr}(L \leq n \leq R) \\
 &= \text{pr}(L \leq 0 \leq R) + 2 \sum_1^{\infty} \{\text{pr}(R \geq n) - \text{pr}(L > n)\} \\
 &= 1 + 2 \sum_1^{\infty} \text{pr}(R \geq n) + o(1) \quad \text{as } \eta \rightarrow \infty \\
 &= 1 + 2n_0 + 2 \sum_{n_0+1}^{\infty} \text{pr}(R \geq n) - 2 \sum_1^{n_0} \{1 - \text{pr}(R \geq n)\} + o(1). \quad (A1)
 \end{aligned}$$

For positive  $n$ , by the definition of  $R$

$$\begin{aligned}
 \text{pr}(R \geq n) &= \text{pr} \left( \sup_{i < n} S_i \leq \sup_{i \geq n} S_i + \eta \right) \\
 &= \int \int_{[-\eta, 0) \times [0, \infty)} \text{pr} \left\{ S_n \in d\xi, \max_{i \geq n} (S_i - S_n) \in dy \right\} \\
 &\quad \times \text{pr} \left( \max_{0 \leq i \leq n} S_i \leq \eta + \xi + y | S_n = \xi \right) \text{pr} \left( \max_{i \leq 0} S_i \leq \eta + \xi + y \right) \\
 &= \int \int_{[-\eta, 0) \times [0, \infty)} \text{pr}(S_n \in d\xi) \text{pr}(M \in dy) \text{pr} \left( \max_{0 \leq i \leq n} S_i \leq \eta + \xi + y | S_n = \xi \right) \text{pr}(M \leq \eta + \xi + y) \\
 &= \int \int_{[0, \infty) \times [0, \infty)} \text{pr}(S_n \in -\eta + dx) \text{pr}(M \in dy) \text{pr} \left( \max_{0 \leq i \leq n} S_i \leq x + y | S_n = x - \eta \right) \text{pr}(M \leq x + y).
 \end{aligned}$$

Let  $n_0 = \lfloor 2\eta \rfloor$  and  $k = n - n_0$ , so

$$\text{pr}(S_n \epsilon - \eta + dx) = \varphi \left\{ \frac{x + k/2 - \eta + n_0/2}{(n_0 + k)^{1/2}} \right\} (n_0 + k)^{-1/2} dx.$$

It may be shown that the contribution to the two series in (A1) from values of  $x$  and  $k$  outside the range  $|k| \leq \eta^{2/3}, |x + k/2| \leq \eta^{2/3}$  is negligible, and inside this range

$$\text{pr} \left( \max_{0 \leq i \leq n_0 + k} S_i > x + y | S_{n_0 + k} = -\eta + x \right) - \text{pr}(M > x + y)$$

converges uniformly to 0. Hence for the purpose of evaluating (A1) asymptotically,  $\text{pr}(R \geq n_0 + k)$  may be replaced by

$$\int_0^{\eta^{2/3}} \int_0^\infty \varphi \left\{ \frac{x + k/2}{(n_0 + k)^{1/2}} \right\} (n_0 + k)^{-1/2} dx \text{pr}(M \epsilon dy) \{1 - 2\text{pr}(M > x + y) + \text{pr}^2(M > x + y)\}.$$

For  $k = 0$  this integral converges to  $1/2$ . The terms in (A1) for  $k = \pm 1, \pm 2, \dots$  may be paired, and after some calculation one obtains

$$\begin{aligned} \sum_{k \geq 1} \text{pr}(R \geq n_0 + k) - \sum_{k \leq 0} \{1 - \text{pr}(R \geq n_0 + k)\} &= -1/2 \\ &+ \sum_{k \geq 1} [\Phi\{2^{-1}k/(n_0 - k)^{1/2}\} - \Phi\{2^{-1}k/(n_0 + k)^{1/2}\}] \\ &- 2n_0^{-1/2} \varphi(2^{-1}/n_0^{1/2}) \int_0^\infty \int_0^\infty \text{pr}(M \epsilon dy) \{2\text{pr}(M > x + y) \\ &- \text{pr}^2(M > x + y)\} dx + o(1). \end{aligned}$$

A Taylor series expansion, approximation of Riemann sums by integrals, and substitution of the result back into (A1) complete the informal proof of Theorem 1.

## APPENDIX B

### Boundary Crossing Probabilities.

This appendix is concerned with approximations to boundary crossing probabilities like (22). The notation used here is independent of the body of the paper.

Let  $x_1, x_2, \dots$  be independent random variables with distribution function of the form

$$dF_\theta(x) = \exp\{\theta x - \psi(\theta)\} dF_0(x) \quad (A2)$$

with  $F_0$  standardized to have mean 0 and variance 1. Let  $S_n = x_1 + \dots + x_n$ . To emphasize dependence of probabilities and expectations on  $\theta$  we write  $\text{pr}_\theta$  and  $E_\theta$ .

Let  $b_0 > 0, b(0) = 0$ , and define

$$T = \inf\{n : S_n > b_0 + mb(n/m)\}.$$

We seek approximations as  $m \rightarrow \infty$  for

$$\text{pr}_0(T < m_0 | S_{m_0} = m_0 \xi_0),$$

where  $m_0 = mt_0$  for some fixed  $t_0 > 0$ . We assume that  $\xi_0 \leq b(t_0)/t_0$  and  $\xi_0 < b(t)/t$  uniformly on compact subsets of  $[0, t_0]$ . Let  $c_1 = b'(0), c_2 = b''(0)/2$ . Note that  $\xi_0 < c_1$  and locally near 0

$$mb(n/m) = c_1 n + c_2 n^2/m + O(n^3/m^2). \quad (A3)$$

Define  $\theta_0$  by  $\psi'(\theta_0) = \xi_0$  and  $\theta_1 > \theta_0$  by

$$\psi(\theta_1) - \psi(\theta_0) = c_1(\theta_1 - \theta_0). \quad (A4)$$

Let  $\xi_1 = \psi'(\theta_1), \sigma_i^2 = \psi''(\theta_i)$  ( $i = 0, 1$ ), and  $\Delta = \theta_1 - \theta_0$ .

Let  $t_+(\eta)(t_-(\eta)) = \inf\{n : S_n - \eta n > (<) 0\}$ , and put

$$\nu_+ = \text{pr}_{\theta_0}\{t_+(c_1) = \infty\} \text{pr}_{\theta_1}\{t_-(c_1) = +\infty\} / \{\Delta(\xi_1 - c_1)\}. \quad (\text{A5})$$

Theorem 3. Assume that for all  $\theta$ , for all sufficiently large  $n$  the  $n$ -fold convolution of  $F_\theta$  has an integrable characteristic function. Suppose  $b_0 \rightarrow \infty$  as  $m \rightarrow \infty$  and  $b_0 = O(\sqrt{m})$ . Then

$$\begin{aligned} & \text{pr}_0(T < m_0 | S_{m_0} = m_0 \xi_0) \\ & \sim \nu_+ \exp[-\Delta b_0 - m^{-1} b_0^2 (\xi_1 - c_1)^{-2} \{(\xi_1 - \xi_0)^2 / (2\sigma_0^2) + \Delta c_2\}]. \end{aligned} \quad (\text{A6})$$

In order to evaluate the constant  $\nu_+$  defined in (A5) it may be helpful to use the local expansion for small  $c_1$  and  $\theta_0$

$$\nu_+ = \exp(-\Delta \rho_+) + o(\Delta^2), \quad (\text{A7})$$

where  $\rho_+ = E_0 S_{t_+(0)}^2 / (2ES_{t_+(0)})$ . See Siegmund (1985, Chapter X) for a justification and method for computing  $\rho_+$  numerically. In the normal case this is the approximation (5). In the exponential case the standardized generating distribution  $dF_0(x)$  equals either  $\exp(-x+1)dx$  ( $x > -1$ ) or  $\exp(x-1)dx$  ( $x < 1$ ), and the approximation (A7) is  $\exp(-\Delta)$  or  $\exp(-\Delta/3)$  respectively. These approximations were used in the numerical example in Section 4.

If  $b_0 = o(\sqrt{m})$ , the right hand side of (A6) is just Cramér's classical approximation to the probability of ultimate ruin, as follows. Suppose  $b(t) \equiv c_1 t$ , and consider  $\text{pr}_{\theta_0}(T < \infty)$ . By (A2) and (A4) the likelihood ratio of  $x_1, \dots, x_n$  under  $\theta_0$  relative to  $\theta_1$  is  $\exp\{-\Delta(S_n - c_1 n)\}$ . Hence by Wald's likelihood ratio identity

$$\text{pr}_{\theta_0}(T < \infty) = \exp(-\Delta b_0) E_{\theta_1} \exp\{-\Delta(S_T - b_0 - c_1 T)\}, \quad (\text{A8})$$

and by the renewal theorem

$$\lim_{b_0 \rightarrow \infty} E_{\theta_1} \exp\{-\Delta(S_T - b_0 - c_1 T)\} = \nu_+,$$

where  $\nu_+$  is defined in (A5). See, for example, Siegmund (1985, Chapter VIII) for details. Hence (A6) has the interpretation that if  $b_0 = o(\sqrt{m})$ , asymptotically the curvature of the boundary plays no role and the conditional probability given  $S_{m_0} = m_0 \xi_0$  is effectively the unconditional probability  $\text{pr}_{\theta_0}$  having the drift  $\xi_0 = \psi'(\theta_0)$ .

In general the first term multiplying  $m^{-1}b_0^2$  on the right hand side of (A6) corrects for the fact that we have a conditional, not an unconditional probability and the second corrects for the curvature in  $b(t)$ . In fact, by (A3) one can modify (A8) to read

$$\begin{aligned} & \exp(\Delta b_0) \text{pr}_{\theta_0}(T < m_0) \\ &= E_{\theta_1}(\exp[-\Delta\{S_T - b_0 - mb(T/m)\} - \Delta c_2 T^2/m + O(T^3/m^2)]; T < m_0). \end{aligned}$$

It is easy to see that  $b_0^{-1}T \rightarrow (\xi_1 - c_1)^{-1}$  in  $\text{pr}_{\theta_1}$ -probability, and by Theorem 9.45 of Siegmund (1985) the limiting distribution of the excess over the boundary,  $S_T - b_0 - mb(T/m)$ , is the same as in the linear case. This explains the correction for non-linearity. The correction to account for the conditional probability is obtained by a modification of the proof of Theorem 8.72 of Siegmund (1985). The details are much more complicated and have been omitted.

To illustrate the approximation (A6) suppose that  $y_1, \dots, y_m$  are independent standard exponential,  $W_n = y_1 + \dots + y_n$ , and consider Pettitt's test with acceptance region (14), or to emphasize the invariance of the exponential scale parameter

$$\left\{ \left| \max_n (nW_m/m - W_n)/(W_m/m) \right| \leq a \right\}.$$

From the theorem one obtains the following approximations:

$$\begin{aligned} & \text{pr} \left\{ \max_{0 < i < m-j} (i - W_{j+i}/x_0) > a - j | W_j = jy_0, W_m = mx_0 \right\} \\ & \cong \exp[-\Delta(b_0 + 1/3) - 2^{-1}b_0^2\Delta^2/\{(m-j)\theta_1^2\}], \end{aligned}$$

where  $b_0 = a - j(1 - y_0/x_0)$ ,  $\xi_0 = -j(1 - y_0/x_0)/(m - j)$ ,  $\theta_0 = \xi_0/(1 - \xi_0) < 0$ , and  $\theta_1 > 0$  satisfies

$$\theta_1 - \log(1 + \theta_1) = \theta_0 - \log(1 + \theta_0);$$

$$\text{pr} \left\{ \max_{0 < i < j} (i - W_i/x_0) > a | W_j = jy_0, W_m = mx_0 \right\} \cong \exp\{-\Delta(b_0 + 1) - 2^{-1}b_0^2\Delta^2/(j\theta_1^2)\},$$

where  $b_0 = a - j(1 - y_0/x_0)$ ,  $\xi_0 = -(1 - y_0/x_0)$ ,  $\theta_0 = \xi_0/(1 + \xi_0) < 0$ , and  $\theta_1 > 0$  satisfies

$$\theta_1 + \log(1 - \theta_1) = \theta_0 + \log(1 - \theta_0).$$

The local expansion (A7) for  $\nu_+$  is used in both these approximations. The analogous approximations for the likelihood ratio test are similar but slightly more complicated since they involve both the first and second derivatives of the boundary curve  $b(t)$  at  $t = jm$ , which are easily obtained from (21).

If we invert the Pettitt test to obtain a 95% confidence set for a change-point in the coal mining accident data, we find the same confidence set as in Section 4, with one exception. The attained significance level of  $j = 129$  is  $.039 + .014 > .05$ , so  $j = 129$ , corresponding to the year 1894 is included in the confidence set.

## References

- Cobb, G. W. (1978). The problem of the Nile: conditional solution to a change point problem, *Biometrika* **62**, 243-51.
- Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* **64**, 247-54.
- Fisher, R. A. (1934). Two new properties of mathematical likelihood, *Proc. R. Soc. A* **144**, 285-307.
- Pollak, M. and Siegmund, D. (1985). A diffusion process and its application to detecting a change in the drift of Brownian motion, *Biometrika* **72**, 267-80.
- Hinkley, D. V. (1970). Inference about the change point in a sequence of random variables, *Biometrika* **57**, 1-17.
- Hinkley, D. V. (1972). Time ordered classification, *Biometrika* **59**, 509-523.
- Hooper, P. M. (1982). Invariant confidence sets with smallest expected measure, *Ann. Statist.* **10**, 1283-94.
- Hora, R. B. and Buehler, R. J. (1966), Fiducial theory and invariant estimation, *Ann. Math. Statist.* **37**, 643-656.
- Jarrett, R. G. (1979). A note on the intervals between coal-mining disasters, *Biometrika* **66**, 191-3.
- Maguire, B. A., Pearson, E. S., and Wynn, A. H. A. (1952). The time intervals between industrial accidents, *Biometrika* **38**, 168-80.
- Pettitt, A. N. (1980). A simple cumulative sum type statistic for the change-point problem with zero-one observations, *Biometrika* **67**, 79-84.
- Raferty, A. E. and Akman, V. E. (1986). Bayesian analysis of a Poisson process with a change-point, *Biometrika* **73**, 85-90.

- Siegmund, D. (1985). *Sequential Analysis: Test and Confidence Intervals*, Springer-Verlag, New York-Heidelberg-Berlin.
- Siegmund, D. (1986). Boundary crossing probabilities and statistical applications, *Ann. Statist.* **14**, 361-404.
- Stein, C. (1965). Approximation of improper prior measures by prior probability measures, *Bernoulli, Bayer, Laplace, Anniversary Volume*, J. Neyman and L. M. Le Cam, eds., Springer-Verlag, New York-Heidelberg-Berlin, 217-240.
- Worsley, K. J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables, *Biometrika* **73**, 91-104.



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 39	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) CONFIDENCE SETS FOR A CHANGE-POINT		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) David Siegmund		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0306
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics - Sequoia Hall Stanford University Stanford, California 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-373
11. CONTROLLING OFFICE NAME AND ADDRESS Statistics & Probability Program Code (411 (SP)) Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE October 1986
		13. NUMBER OF PAGES 28pp.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)  Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Several methods are discussed for confidence set estimation of a change-point in a sequence of independent observations from completely specified distributions. The method based on the likelihood ratio statistic is extended to the case of independent observations from a one parameter exponential family. Joint confidence sets for the change-point and the parameters of the exponential family are also considered.		

END

12-86

DTIC